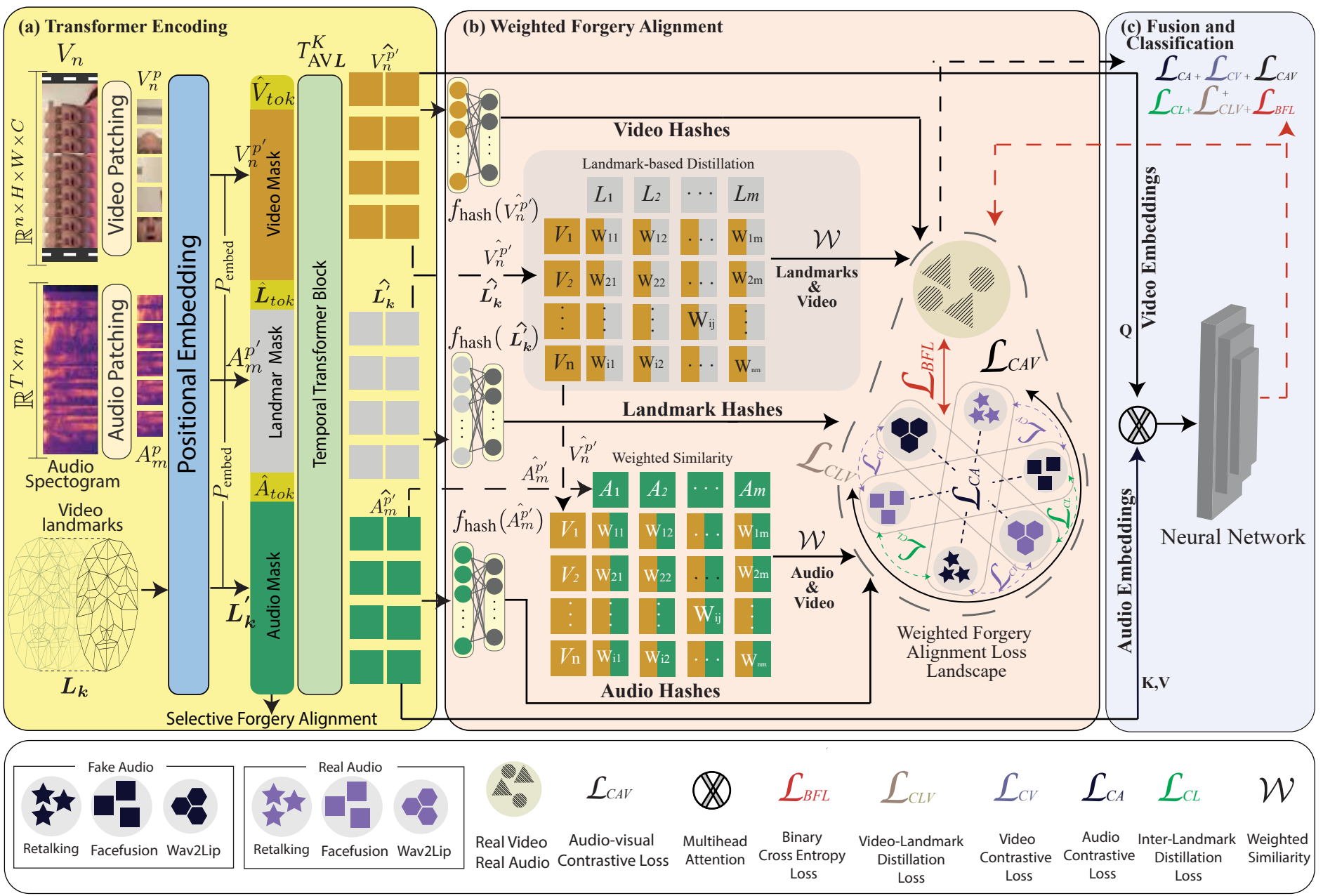
(a) Transformer Encoding
(b) Weighted Forgery Alignment
(c) Fusion and Classification
Video Patching
Audio Patching
Audio Spectogram
Video landmarks
Positional Embedding
Video Mask
Landmar Mask
Audio Mask
Temporal Transformer Block
Selective Forgery Alignment
Video Hashes
Landmark-based Distillation
Landmarks & Video
Landmark Hashes
Weighted Similarity
Audio & Video
Audio Hashes
Weighted Forgery Alignment Loss Landscape
Video Embeddings
Audio Embeddings
Q
K,V
Neural Network
Fake Audio
Retalking
Facefusion
Wav2Lip
Real Audio
Retalking
Facefusion
Wav2Lip
Real Video Real Audio
Audio-visual Contrastive Loss
Multihead Attention
Binary Cross Entropy Loss
Video-Landmark Distillation Loss
Video Contrastive Loss
Audio Contrastive Loss
Inter-Landmark Distillation Loss
Weighted Similiarity